Sentiue: Target and Aspect based Sentiment Analysis in SemEval-2015 Task 12

José Saias
DI - ECT - Universidade de Évora
Rua Romão Ramalho, 59
7000-671 Évora, Portugal
`jsaias@uevora.pt`

Abstract

This paper describes our participation in SemEval-2015 Task 12, and the opinion mining system `sentiue`. The general idea is that systems must determine the polarity of the sentiment expressed about a certain aspect of a target entity. For slot 1, entity and attribute category detection, our system applies a supervised machine learning classifier, for each label, followed by a selection based on the probability of the entity/attribute pair, on that domain. The target expression detection, for slot 2, is achieved by using a catalog of known targets for each entity type, complemented with named entity recognition. In the opinion sentiment slot, we used a 3 class polarity classifier, having BoW, lemmas, bigrams after verbs, presence of polarized terms, and punctuation based features. Working in unconstrained mode, our results for slot 1 were assessed with precision between 57% and 63%, and recall varying between 42% and 47%. In sentiment polarity, `sentiue`'s result accuracy was approximately 79%, reaching the best score in 2 of the 3 domains.

1 Introduction

Social networks and other online platforms are an important communication mechanism in current lifestyle. These platforms aggregate user-generated content, such as opinions that people write and publish freely on the Web, and are now valued for market research and trend analysis. Natural language processing (NLP) helps to automatically extract information from these written opinions.

This paper describes a participation in SemEval-2015 Task 12[1], Aspect Based Sentiment Analysis (Pontiki et al., 2015), with the `sentiue` system, from Universidade de Évora. In previous editions of SemEval, we participated in Sentiment Analysis (SA) tasks, but in terms of overall polarity, over Twitter messages (Rosenthal et al., 2014), not being aspect oriented. The general idea for this challenge, is that, for a text, the system must determine the polarity of the sentiment expressed about a certain aspect of a particular target entity. Our `sentiue` system is an evolution from our previous work (Saias and Fernandes, 2013; Saias, 2014), for target oriented SA.

Task 12 was run in two phases. In phase A systems are tested for aspect detection with one slot to aspect category, and a second slot for the opinion target expression on the text. Test data includes review texts for two domains: restaurants and laptops. In phase B, aspect category is provided, and systems must assign a polarity (positive, negative, or neutral) for each opinion. In this phase, systems received also texts from a third domain, hotels, for which no sentiment training data was given.

We used a supervised machine learning classifier combined with a probability based selection process, for entity and attribute category detection, on slot 1. Target expression detection was performed with an entity catalog, filled with known targets for each entity type, and named entity recognition (NER). For the sentiment polarity slot, we used a a supervised machine learning classifier, having bag-of-words (BoW), lemmas, bigrams after verbs, and

[1]http://alt.qcri.org/semeval2015/task12/

punctuation based features, along with sentiment lexicon based features. The detailed procedure is explained in section 3.

2 Related Work

Many SA related publications, originating both in industry and in academia, have appeared, and it is notorious the growing interest by companies. Popular scientific forums and events include activities and workshops on this area, such as RepLab (Amigó et al., 2014) at CLEF[2], for online reputation, or ABSA and Twitter SA tasks in SemEval.

In last year's edition of this SemEval task (Pontiki et al., 2014), there were 26 systems participating in the polarity subtask. The two systems with better polarity classification accuracy were from NRC-Canada and DCU teams. NRC-Canada system (Kiritchenko et al., 2014) was trained with the data provided in the task, and complemented with lexicons generated from other corpora of customer reviews, to help feature extraction in machine learning. Stanford CoreNLP was used to tokenize, POS tagging, and dependency parse trees. They address polarity classification with a linear SVM classifier, with features for: the target, and its surrounding words; POS based features; dependency tree based features; unigrams and bigrams; lexicon based features. The DCU system (Wagner et al., 2014) also uses SVM for aspect and for polarity classification, combining bag-of-n-gram features with rule-based features. N-grams (with size from 1 to 5) in a window around the aspect term, are used as features, as well as features derived from a sentiment lexicon. The rule-based approach to predict the polarity of an aspect term, generated features considering all words score and their distance to the aspect term.

3 Method

Our participation involved the adaptation of our previous real-time system, for text overall sentiment classification, into a target oriented SA system. The next subsections explain how the system works, for each part of Task 12 challenge.

3.1 Aspect Entity and Attribute

The first annotation task focuses on aspect category. This category is an **entity** and **attribute** pair, each chosen from an inventory with possible values, in each domain, for entity types and attributes. Since the possible category types are known and limited, we decided to use a classifier for each entity type (e.g. *food, laptop*) and for each attribute label (e.g. *price, quality*). Our approach comprises two stages. The first processes each review sentence assigning to it zero, one, or more entity types and attribute labels. The second stage chooses and combines identified entities and attributes, forming the aspect annotation. Analyzing the training data, we found that in the same sentence, there may be opinions on various types of entity (e.g. CPU, battery) or attributes. Thus, we have chosen to train a classifier for each entity type, and a classifier for each attribute label. We set a supervised machine learning text classifier, using MALLET (McCallum, 2002), a Java-based tool for NLP, with machine learning applications to text. For the purpose of this stage, it was necessary to prepare the training data for each binary classifier, that would determine whether a sentence contains an opinion on its tag (entity type or attribute label). The train process was the same for all tags, entity type or attribute label, of each domain. We created a dataset where each instance is a sentence text, and its class is *tag*, if the sentence had at least one opinion with that tag, or *no_tag* otherwise. Text preprocessing includes tokenization, POS tagging and lemmatization, all performed with Stanford CoreNLP (Toutanova et al., 2003; Manning et al., 2014) tool. The classifier algorithm was Maximum Entropy[3], and the classifier model features were text words and lemmas.

Second stage starts with each sentence annotated with a set and tags, some for entity type and some for attribute label. When a sentence has no annotations, the system assumes that there is no opinion. In case of 1 tag on entity type and 1 tag in the attribute label, then it is the trivial case where the junction of the two results in the aspect annotation. For sentences with 1 tag on entity type and 0 tags for the attribute, our system searches for the most frequent aspect annotation, within the sentence domain, that

---

includes that entity type. The equivalent is applied in the case of 0 tags to entity and 1 tag for the attribute label. If both sides have one or more tags, the system applies a cycle, where each loop iteration forms the more frequent pair (entity,attribute) in that domain, and removes these two tags from the sentence tag set. This is repeated until the first, entity or attribute side, exhausts the tags provided by the previous stage classifier. And if some tags are left, on the opposite side of the pair, the system applies, for each, the same process already explained for case 0-1 or 1-0.

3.2 Opinion Target Expression

At this point, sentences are already marked as having (or not) opinions on certain aspect category. For each opinion on restaurants domain, the system needed to identify the entity mention on the sentence text, referred to as the opinion target expression (OTE). We collected the opinion targets for each entity type, from the training data, forming a catalog. If any of the targets already known (e.g. restaurant name, or meal) appears in the sentence text, next to a verb or adjective, it is chosen as the OTE. If this does not lead to any OTE candidate, our system applies named entity recognition, looking for references to organization and location entities, using Stanford NER tool (Finkel et al., 2005; Manning et al., 2014). Having found one OTE, through the catalog or by NER, its text and position are marked in slot 2. If no mention is found, OTE slot is filled with the NULL value.

3.3 Sentiment Polarity

Phase B was held in a subsequent period, and the input given to the systems is a little different, having the correct annotations on the aspect category, in restaurants, laptops and hotels domains. For each opinion, the participating systems must assign a sentiment polarity (positive, negative or neutral), considering the opinion aspect.

For training, there were 1654 opinions on restaurants domain, and 1974 more opinions about laptops, all annotated for polarity. No sentiment training data was given for hotels domain. Considering the available data, and the objective of this phase, we used a supervised machine learning classifier to predict each opinion polarity. Instead of multiple classi-

fiers, such as implemented for slot 1, we prepared a single classifier, thought, as before, for text but tuned with a different model, so that it can choose between positive, negative or neutral polarity.

Sentences without opinion are not considered in the training, because here the polarity is associated with opinions. Further, a single sentence may have several opinions about different aspects, and each may have a different and independent polarity. To train the classifier, for each opinion we created a polarity data instance, containing the sentence text, its domain, its aspect entity and attribute, OTE (if available, in restaurants), and the opinion polarity to be learned. As before, MALLET was used with a Maximum Entropy classifier. The sentence text preprocessing was the same we did for aspect category classification. The features to represent each instance were:

- BoW with a feature for each token text;

- lemmas for verbs and adjectives;

- bigram after verb (lemmatized);

- presence of negation terms;

- bigram after negation term;

- presence of exclamation/question mark;

- presence of polarized terms (positive or negative), according to each sentiment lexicon;

- whether there are polarized terms before exclamation mark and question mark;

- bigram before, and after, any polarized term;

- polarity inversion, by negation detection before some polarized term;

- presence of polarized terms in the last 5 tokens;

- a feature for the domain, and two features for the entity type and the attribute label.

To see whether a term is polarized, each token text is verified in each sentiment lexicon. These polarity support resources are AFINN lexicon (Nielsen, 2011), Bing Liu's opinion lexicon (Liu et al., 2005) and MPQA subjectivity clues (Wiebe et al., 2005).

Domain	Precision	Recall	F-measure
restaurants	0,633	0,472	0,541
laptops	0,577	0,441	0,500

Table 1: `sentiue`'s evaluation on aspect category.

Domain	Precision	Recall	F-measure
restaurants	0,488	0,336	0,398

Table 2: `sentiue`'s evaluation on target detection.

Domain	Accuracy
restaurants	0,787
laptops	0,793
hotels	0,788

Table 3: `sentiue` accuracy on sentiment polarity.

Domain,Polarity	Precision	Recall	F-measure
restaurants, positive	0,767	0,914	0,834
restaurants, negative	0,825	0,708	0,762
restaurants, neutral	0,714	0,111	0,192
laptops, positive	0,831	0,891	0,860
laptops, negative	0,766	0,787	0,777
laptops, neutral	0,387	0,152	0,218
hotels, positive	0,887	0,840	0,863
hotels, negative	0,608	0,738	0,667
hotels, neutral	0,143	0,083	0,105

Table 4: `sentiue`'s SA evaluation per domain.

After some experimentation, we decided to use a single full train, joining the instances of restaurants and laptops as a whole training set. The resulting model was used to classify the opinion polarity for the three domains.

Because we used sentiment lexicons, our system operates in unconstrained mode. These additional resources served as support for features extraction. No supplementary training texts were used. In our development testing, we obtained an 80% accuracy for polarity. After this, the result is written in XML format for submission.

4   Results

The phase A test data had 685 sentences on restaurants domain and 761 on laptops domain. With the method described above, the `sentiue` system extracted 596 opinion categories for restaurants domain and 751 other for laptops domain.

Table 1 shows the evaluation for slot 1. Among the 15 submissions evaluated in the first domain, the best system F-score value was 0,627, while our result F-score was 0,541. For laptops aspect category, `sentiue`'s scores were lower, but improving in the comparison with other systems, achieving the second best F-measure, out of 9 evaluated submissions. The evaluation of our result in opinion target expression in given in Table 2. This was a poor result, when compared with the 0,524 average F-measure of the 21 submissions for this slot.

In phase B systems had to fill the slot 3, with sentiment polarity to 845 opinions on restaurants domain, 949 opinions on laptops domain, and 339 opinions on hotels domain.

On Table 3 we find our system's result accuracy, in the two trained domains plus the untrained hotels do-

main. In this slot we got the most satisfactory result, with the best accuracy in restaurants and laptops, and an above average score, in the hotels domain. The detailed evaluation is shown in Table 4, with values for precision, recall and f-measure, per domain and polarity class.

5   Conclusions

By participating in this SemEval edition, we sought to develop our previous work, in order to achieve SA results focused on the opinion targets.

Our results were poor for OTE detection, but we think it will be easy to correct the implementation problems for that part. As example, while checking if a sentence contained a known target, from the catalog, the system did not require whole words to be matched, and this led to some misidentification of word substrings as target.

Our result was more satisfactory for slot 1, with a F-measure slightly above average between the 15 evaluated submissions for restaurants domain, and 4.5% better than submissions average for laptops domain. The distribution of opinions for each aspect category is not uniform. For example, for attribute label classification, we already know that QUALITY and GENERAL have much more instances than other labels. This analysis inspired our approach in the second stage, explained in section 3.1. To improve this part, we think to introduce a cascade classifier.

After the classification obtained in the current first stage, other machine learning classifier will decide how to pair entity+attribute, based on the wording of the sentence. Another future work idea is to use more corpora for training the aspect classifiers, as other systems (Kiritchenko et al., 2014) have tried.

In phase B `sentiue` achieved good results. This, perhaps, is justified by our previous experience in overall SA. Many of the polarity classifier features are inherited from our former system. SemEval challenge is always a motivation to test our system and an opportunity to learn from other participants.

Acknowledgments

References

Enrique Amigó, Jorge Carrillo-de-Albornoz, Irina Chugur, Adolfo Corujo, Julio Gonzalo, Edgar Meij, Maarten de Rijke, Damiano Spina. 2014. Overview of RepLab 2014: Author Profiling and Reputation Dimensions for Online Reputation Management. In Information Access Evaluation. Multilinguality, Multimodality, and Interaction. Lecture Notes in Computer Science, Volume 8685, 2014, pp 307-322.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363-370.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif M. Mohammad. 2014. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). p. 437–442. Dublin, Ireland, August 2014.

Bing Liu, Minqing Hu and Junsheng Cheng. 2005. *Opinion Observer: Analyzing and Comparing Opinions on the Web*. In Proceedings of the 14th International World Wide Web conference (WWW-2005). Chiba, Japan.

Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

Andrew Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit.

Finn Årup Nielsen. 2011. *A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs*. In Proceedings, 1st Workshop on Making Sense of Microposts (#MSM2011): Big things come in small packages. pp: 93-98. Greece.

Maria Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. Proceedings of the 8th SemEval, 2014. Dublin, Ireland.

Maria Pontiki, Dimitrios Galanis, Haris Papageogiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado.

Sara Rosenthal, Alan Ritter, Veselin Stoyanov, and Preslav Nakov. 2014. SemEval-2014 Task 9: Sentiment Analysis in Twitter. In Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval'14). August 23-24, 2014, Dublin, Ireland.

José Saias and Hilário Fernandes. 2013. Senti.ue-en: An approach for informally written short texts in SemEval-2013 Sentiment Analysis task. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 508–512, Atlanta, Georgia, USA.

José Saias. Senti.ue: Tweet overall sentiment classification approach for SemEval-2014 task 9. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 546–550, Dublin, Ireland, August 2014. ISBN 978-1-941643-24-2.

Kristina Toutanova, Dan Klein, Christopher Manning, and Yoram Singer. 2003. Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In Proceedings of HLT-NAACL 2003, pp. 252-259.

Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster and Lamia Tounsi. 2014. DCU: Aspect-based Polarity Classification for SemEval Task 4. Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). p. 223–229. Dublin, Ireland, August 2014.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. Language Resources and Evaluation, 39:165–210.